Introducing the Rest of Government to openSAFELY

A new secure analytics platform, "openSAFELY.org", has emerged from the health response to COVID-19.

The openSAFELY platform allows reliable and repeatable analysis of sensitive, i.e. GDPR special category, personal data while remaining under the control and guardianship of large numbers of individual data controllers[1] who share a small number of data processors in common,[2] *without needing to create a new copy of the data*. (The step of copying the data being the one that usually causes problems.)

The openSAFELY approach removes the significant risk of data copies being misplaced, compromised, misused or abused, as it involves no copying of data – patients' data is processed by existing data processors, acting on behalf of their existing data controllers.

This approach only works when there is a **clear governance framework** for data originating within many data controllers, all of which are served by a small number of data processors.

In the NHS, 8,000 GPs are served by four data processors,[3] the "GP Systems of Choice" Electronic Health Record (EHR) providers. In the Department for Education, of 24,000 schools, 80% are served by one data processor, Capita, with its School Information Management System (SIMS), while another dozen or so companies service the remaining 20%. In the Ministry of Housing, Communities & Local Government, 350 councils are served by a tiny number of data processors for particular functions with, e.g. only three main players providing Integrated Children's Systems for child services.[4]

Beyond its application in health, openSAFELY provides a specification and tooling for data processors to safely produce the statistics researchers choose – where approved by the appropriate authorities, and subject to statistical protections and any/all other governance requirements, including data protection mandated dissent – while allowing data controllers to retain their controllership.

By remaining **consensual, safe and transparent** throughout, data processors are able to meet the requirements of both data controllers and researchers; the openSAFELY infrastructure combines the results of queries from each data processor to reconstitute national statistics, and returns only that 'safe' output to the requestor.

All of the stakeholders involved have their legal rights and responsibilities maintained, with good governance and transparency embedded throughout the process, making better good practice the norm. Fundamentally, *no data is copied* – which is what usually lies at the root of concerns around control.

---

[1] i.e. GP practices
[2] i.e. the GPSoC IT providers, in this case TPP.
[3] In practice, three providers (EMIS, TPP and INPS) provide 99%+ of EHR services to GPs in England.
[4] A 2018 Ofsted survey of 79 LAs found the main providers of ICS IT to be LiquidLogic and CoreLogic: https://socialcareinspection.blog.gov.uk/2018/10/11/integrated-childrens-systems-what-local-authorities-have-told-us/ noting that "at least one provider [Capita] was not referenced in the responses and that the market share of one of the companies was far greater than this survey indicates."

As demonstrated by openSAFELY's publications,[5] national statistics and analyses can be produced safely, respecting data governance, without data being copied or any high risk disseminations:

- Data controllers can continue to act as responsible data controllers; and
- Data processors get to act as responsible data processors, facilitating the wishes of their data controllers (and data subjects) who wish to support both high quality research and public confidence in data use; and
- Researchers or statisticians can get answers to questions they can legitimately ask, with clear and known standard recodes of complex encodings to simple norms; and
- National bodies get to maintain ongoing statistics access and long-term research.

In practice (DfE)

Given long-held concerns about government re-use of live School Census data, openSAFELY could replace the termly school census with an openSAFELY real-time statistical reporting system that provides current answers for (pre)specified queries.

There is no national copy of current data which can be compelled by any other Government Department. And when a pupil leaves a school, their data would be extracted (as now) to the central archive of 'non-current' pupils.

The openSAFELY approach would make it possible for data controllers to restrict queries to the data they control, and to delegate that restriction (as is currently done for access to NPD) in a normalised way, while transparently showing what is being processed – making it impossible to target individuals, and ensuring that every query can be audited.

Researchers would gain a clearer understanding of the coding and data that is available and used, and could also have a greater degree of confidence in the nuances of the data.

The openSAFELY approach would allow DfE to produce better statistics more quickly and efficiently; provide a much higher quality service to researchers; offer far better data query tools internally, within a clear and transparent governance framework; and mitigate many of the controversies around NPD over recent years as NPD would not be a contemporary dataset, through a transition to a fundamentally different school census collection model that has been proven with special category personal data in the NHS during COVID. The work of the various 'information intermediaries' for the school census could be automated away by using infrastructure openSAFELY has built for health data, which can be repurposed.

Avoiding the mistakes of ContactPoint

ContactPoint was decommissioned ultimately because the data copying issues proved entirely insurmountable. openSAFELY is a mechanism to allow national analysis of locally-held data *without* data copies – there is no data copying involved. That Capita operates as the data processor, and often sole supplier, to local authority data controllers means there will be in many

[5] e.g. https://opensafely.org/outputs/2020/05/covid-risk-factors/

cases a single place to change the process, and to resolve many of the outstanding ‘wicked problems’ around policy, data protection, and politics.

In particular, for those issues that ContactPoint was intended to address, openSAFELY would allow a (confidential) answer to the (sensitive and confidential) question, “In which councils has a child with these specified characteristics appeared?” – thereby allowing relevant staff to have the appropriate conversations across LA boundaries, without a central database and the panoply of attendant risks that brings.

Such deeply sensitive queries – banned under most normal Information Governance – were the main goal of ContactPoint. And that such queries are possible highlights the need for ongoing IG across all openSAFELY implementations.